基于改进的深度残差网络的表情识别研究 *

何　俊，刘　跃，李倡洪，沈津铭，李　帅，王京威
(南昌大学 信息工程学院，南昌 330031)

摘　要：提出了一种基于改进的深度残差网络（residual network，ResNet）的表情识别算法。采用小卷积核和深网络结构，利用残差模块学习残差映射解决了随着网络深度的增加网络精度下降问题，通过迁移学习方法克服了因数据量不足导致训练不充分的缺点；网络架构使用了线性支持向量机（SVM）进行分类。实验中首先利用 ImageNet 数据库进行网络参数预训练，使网络具有良好的提取特征能力，根据迁移学习方法，利用 FER-2013 数据库以及扩充后的 CK+数据库进行参数微调和训练。该算法克服了浅层网络需要依靠手工特征，深层网络难以训练等问题，在 CK+数据库以及 GENKI-4K 数据库上分别取得了 91.333%和 95.775%识别率。SVM 在 CK+数据库的分类效果较 Softmax 提高了 1%左右。

关键词：深度学习；残差网络；表情识别；迁移学习；支持向量机
中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.10.0846

Research on expression recognition based on improved deep residual network

He Jun, Liu Yue, Li Changhong, Shen Jinming, Li Shuai, Wang Jingwei
(*School of Information Engineering*, *Nanchang University*, *Nanchang* 330031, *China*)

Abstract: This paper proposed an improved residual network (ResNet) expression recognition algorithm. The algorithm used small convolution kernels and a deep network structure to solve the problem of accuracy reduction with the increase of depth by the residual module. The experiment overcomes the shortcoming of insufficient data through transfer learning, which can effectively prevent overfitting. The network architecture uses a linear support vector machine (SVM) for classification. The experiment used the ImageNet database to pre-train network parameters to have an excellent ability to extract feature. According to transfer learning, the algorithm used the FER-2013 database and the expanded CK+ database to fine-tune and train network parameters, and overcame the problem that shallow networks rely on manual features and deep networks are difficult to train. The results show the recognition rates is 91.333% and 95.775% on the CK+ database and the GENKI-4K database, respectively. The classification accuracy of SVM in CK + database is about 1% higher than that of Softmax.

Key words: deep learning; residual network; facial expression recognition; transfer learning; support vector machine

0　引言

传统的表情特征提取方法大多依赖手工设计的特征，如 LBP、SIFT 等，不仅设计困难、无法保证这些特征的最优性，而且无法提取图像的高阶统计特征。于是研究者们开始利用深度学习进行表情识别。目前深度神经网络已经被证明在图像、语音、文本领域具有挖掘数据深层潜在的分布式表达特征的能力。其中，卷积神经网络（CNN）用于识别位移、缩放及其他形式扭曲不变性的二维图形的效果尤为突出并广泛应用于图像识别与分类领域[1~4]。CNN 的特征提取层通过训练数据隐式地进行学习，避免了显示的特征抽取。但想要利用深度卷积神经网络完成面部表情识别任务，需要大量的训练数据来训练模型的参数。由于表情识别数据库远远不能满足网络参数训练的要求，所以目前深度学习表情识别算法多是深度学习网络和表情特征相结合的方法。深度学习网络的深度对最后的分类和识别的效果有着很大的影响。研究发现随着网络深度的增加,系统精度得到饱和之后，梯度消失的现象就越来越明显,精度迅速的下滑。但是浅层网络又无法明显提升网络的识别效果[2]。

2015 年，何凯明等人提出了残差网络 [5],该网络不仅解决了这个问题，而且较其他模型识别效果更优秀。受此启发，本文提出一种基于残差网络的改进算法，通过深度残差网络与支持向量机(support vector machine, SVM)[6~9]相结合实现对人脸表情的识别,该算法不依赖任何表情特征，通过深层网络提取表情的深层特征。实验中使用 ResNet-50 模型，实验中利用数据增广技术以及迁移学习使网络得到充分的训练。首先利用经 MTCNN(Multitask Cascaded Convolutional Networks）[10]对数据进行预处理，利用非表情数据库对网络进行预训练，然后利用表情数据进行网络的参数微调，得到最终的模型参数。为了测试网络的可行性，本文对比实验了 InceptionV4，VGG 以及 ResNet+Softmax，经测试训练后，本文算法均优于其他网络和传统方法的效果。

收稿日期：2018-10-15；**修回日期：**2018-12-12　**基金项目：**国家自然科学基金资助项目（61463034）

作者简介：何俊（1969-），男，江西东乡人，教授，博士，硕导，主要研究方向为数据挖掘、人机交互技术、模式识别等（boxhejun@tom.com）；刘跃（1992-），男，安徽蒙城人，硕士研究生，主要研究方向为深度学习、模式识别、人机交互；李倡洪（1962-），男，讲师，主要研究方向为模式识别. 沈津铭（1995-），男，硕士研究生，主要研究方向为深度学习，表情识别；李帅（1994-），男，硕士研究生，主要研究方向为深度学习，模式识别；王京威（1993-），男，硕士研究生，主要研究方向为模式识别与智能控制.

1　残差网络设计

ResNet 较传统卷积神经网络而言引入了残差块结构，也就是在标准的前馈卷积网络上，每两层或三层一个跳跃连接，每个跳跃连接就产生一个残差块，卷积层预测加上输入张量的残差。网络结构中使用 3×3、1×1 的小卷积核,网络更深。1×1 的滤波器减少特征映射的个数，可以有效地减少网络参数，提高网络学习速率。网络结构如图 1 所示，其中 Softmax 为 ResNet-50 分类器，SVM 为改进残差网络分类器。

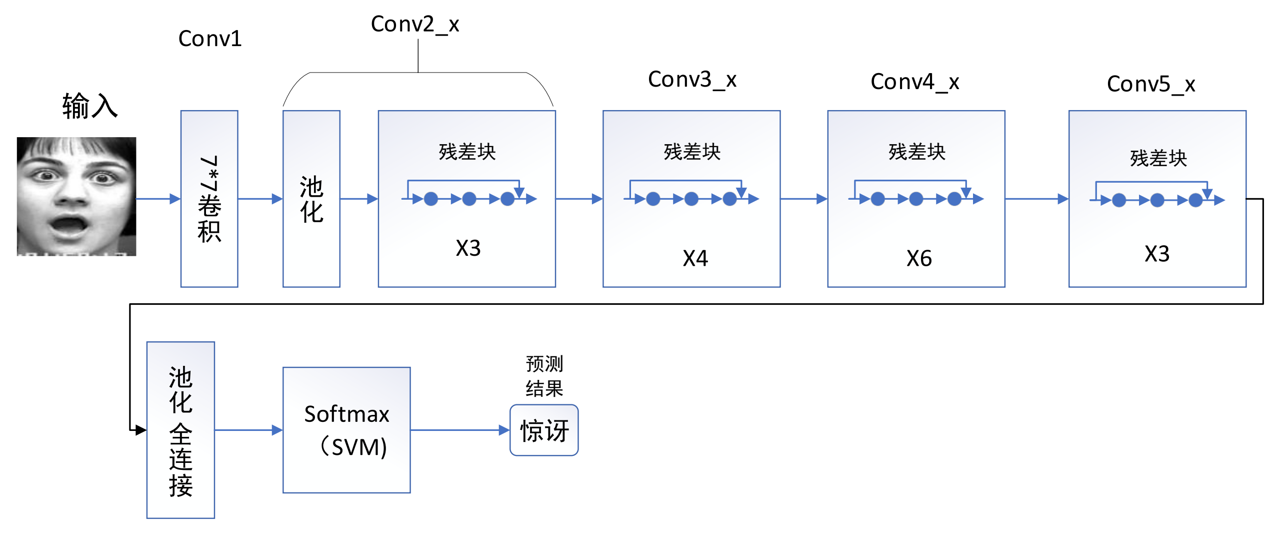


图 1　ResNet-50 结构图

Fig. 1　Structure of ResNet-50

1.1　残差块

残差块结构如图 2 所示。

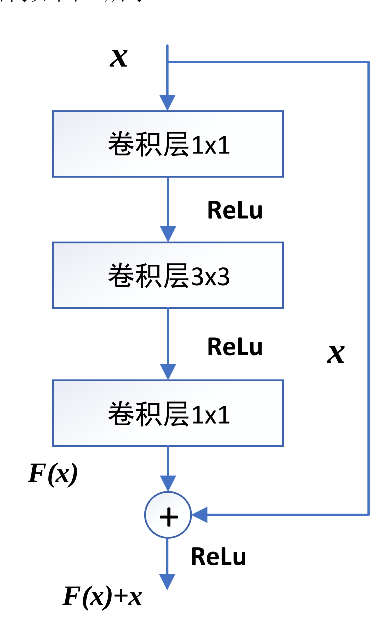


图 2 残差块结构图

Fig. 2　Structure of residual block

残差块跳跃结构增加一个恒等映射即 x，经过卷积层运算后输出为 F（x），激活函数采用 ReLu，ReLU 的使用缓解了网络深度带来的梯度发散问题，促进了梯度在反向传播中路径中的流动。因此，在训练网络时利用 ReLU 可以有效提高训练速度。将 H（x）假设为由几个堆叠层（残差网络采用三个堆叠层）匹配的基础映射，用 x 表示这些第一层的输入。将原始所需要学的函数 H(x)转换成 F(x)+ x。即输出

$$H(x)=x+F(x) \tag{1}$$

可以推广到由浅层 l 到深层 L 的学习特征为

$$H(x_L)=x_l+\sum_{i=l}^{L-1}F(x_i) \tag{2}$$

求得反向过程梯度为

$$\frac{\partial loss}{\partial x_l}=\frac{\partial loss}{\partial H(x_L)}\times\frac{\partial H(x_L)}{\partial x_l} \tag{3}$$

$$\frac{\partial loss}{\partial x_l}=\frac{\partial loss}{\partial H(x_L)}\times(1+\frac{\partial}{\partial x_l}\sum_{l}^{L-1}F(x_i)) \tag{4}$$

$\frac{\partial loss}{\partial x_l}$ 表示损失函数到达的梯度，由式（4）可以看出残差梯度需要经过带有权重的层，而不是直接传递过来，有 1 的存在也不会导致梯度消失，所以残差学习会更容易。

从式(1)可以看出如果前面层已经达到一个最优的函数，那下一层就是没有必要的了，ResNet 通过这种跳跃结构，将优化目标从一个等价映射变为逼近零了，使得优化问题变得很简单。通过这种方式就可以解决网络太深难训练的问题。残差网络使得前馈式/反向传播算法非常顺利进行，使得优化较深层模型更为简单。需要指出的是这个残差块往往需要两层以上，单单一层的残差块并不能起到提升作用。

当输入与输出的维度一样时，无须做其他处理，两者相加即可，但当两者维度不同时，输入要进行变换以后去匹配输出的维度，主要经过两种方式：a）用 zero-padding 去增加维度，此时一般要先做一个下采样，这样不会增加参数；b）用 1×1 卷积来增加维度,这样会增加参数,也会增加计算量。

1.2　设计原则

ResNet 主要是受 VGG 网络[11]启发，主要采用 3×3,1×1 滤波器，遵循两个设计原则：a）对于相同输出特征图尺寸，卷积层有相同个数的滤波器;b）如果特征图尺寸缩小一半，滤波器个数加倍以保持每个层的计算复杂度。在遵循在以上的设计原则的基础上，本文增加了“跳跃连接”。需要指出，这个网络与 VGG 相比，滤波器要少/复杂度要小。

1.3　分类器设计

大多数的深度学习方法使用 Softmax 来进行分类。SVM 分类器作为一种具有较强泛化能力的通用学习算法，被广泛应用于图像识别领域并取得良好的效果。由于 SVM 分类器对大数据高维特征的分类支持较好，本文使用 L2-SVM 的目

标训练深度神经网络进行分类。通过反向传播来自顶层线性 SVM 的梯度来学习较低层权重。为了验证本文算法的有效性，选择 SVM 分类器和 Softmax 分类器对进行对比实验。

本实验采用 LIBSVM 工具[12]实现与残差网络的连接，实现表情的 7 分类。LIBSVM 是基于支持向量机实现的开源库，主要用于分类(支持二分类和多分类)和回归。支持 C++、Java、MATLAB、Python 等多种开发语言 LIBSVM 具有操作简单、易于使用、快速有效、且对 SVM 所涉及的参数调节相对较少的特点。

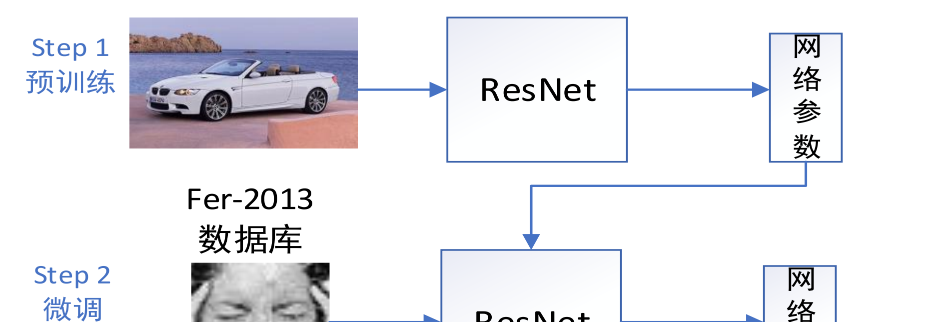
2 基于 ResNet 的表情识别

由于模型参数较多，为满足模型参数训练的需要，引入迁移学习。本文主要选用四个数据库：大量图片数据的 ImageNet 数据库[13]（预训练）、FER-2013[14]（训练）以及 CK+ (extended Cohn Kanade dataset)数据库[15]（训练测试）、GENKI-4K[16]（测试）。

为了图片数据更好的适应 ResNet，提高识别效率，图片均经过 MTCNN 算法进行人脸裁剪，并归一化为 224×224。实验中总共训练七种表情，即生气、厌恶、恐惧、开心、伤心、惊讶和中性。网络训练总共分为两个阶段（图 3）：a)利用 Imagenet 数据库进行参数预训练;b)通过迁移算法 FER-2013 数据库和 CK+库进行参数微调。在经过上述两个训练阶段后，对残差网络进行性能测试。

2.1 数据库

ImageNet 数据集是目前深度学习图像领域应用得非常多的一个领域，关于图像分类、定位、检测等研究工作大多基于此数据集展开。Imagenet 数据集拥有超过 1400 万幅图片，涵盖 20000 多个类别；其中有超过百万的图片有明确的类别标注和图像中物体位置的标注。它广泛应用于计算机视觉领域的研究论文中，几乎成为了深度学习图像领域算法性能检验的“标准”数据集。本文利用该数据训练网络参数，使其具有良好提取特征能力。



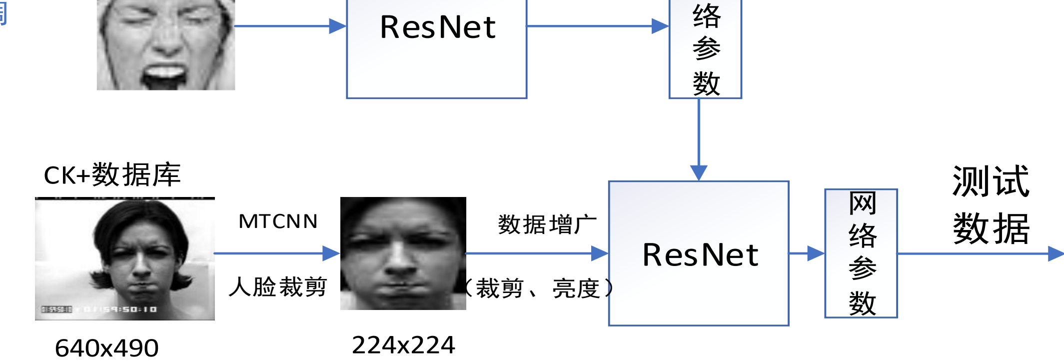


图 3 ResNet-50 训练流程图

Fig. 3 The training process of resnet-50

FER-2013 数据库是 Kaggle 人脸表情分析比赛提供的一个数据集。数据库中包含 35887 张带表情标签的图片，包含生气、厌恶、恐惧、高兴、悲伤、惊讶和正常七种类别的图像，数据库图片大多来自网络，其中包含人脸角度，光照环境等,并且很多图像都有手、头发和围巾等遮挡物的遮挡。

CK+表情数据库是包含的人脸表情是由 123 人共 593 个由自然到高峰的表情序列。其中总共也包含了 8 种基本的表情：生气、蔑视、高兴、悲伤、惊奇、讨厌、害怕、中性。本实验只选择其中七种表情进行识别。选择每个有标签的表情序列 3 至 5 张作为数据，其中选取训练（测试）图片生气 200（20）张，厌恶 250（40）张、恐惧 150（20）张、开心 300（45）张、伤心 150（20）张、惊讶 350（35）张和中性 260（30）张，总共 1600（210）张。，实验在 CK+数据库选用表情使用 MTCNN 对数据库进行人脸裁剪，然后利用数据增广技术（翻转、亮度调节等操作）扩充数据至原数据的 12 倍，此外对每张训练数据（测试图片不使用此操作）进行十字切割，即 25600（2268）张图片,如图 4 所示。

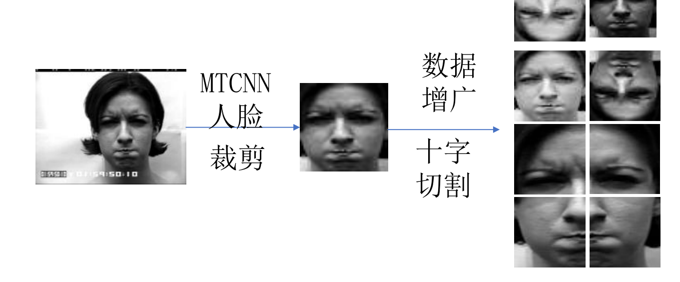


图 4 CK+数据库预处理

Fig. 4 CK+ database preprocessing

使用公开数据库 GENKI-4K 来进行模型泛化能力测试，该数据库包含 4 000 张人脸照片，这些照片存在了各种复杂的变化，包括年龄、肤色、种族、姿势、光照和环境等。在这些图片中，有 2 162 张照片被标注为微笑，而有 1 838 张照片被标注为非微笑。不同于其他一些在实验室中采集的数据集，该数据集可以很好的反映在现实生活中遇到的各种各

样的具有挑战性的笑脸识别问题。

2.2 迁移学习

迁移学习(transfer learning)旨在使用源领域中的知识去改进对于目标领域的预测函数[17]，其核心是将不同领域的知识共享。表 1 对比介绍了不同的迁移学习方法和传统机器学习方法。

表 1　迁移学习与传统机器学习

Table 1　Transfer Learning and Traditional Machine Learning

领域		源领域与目标域	源任务与目标任务
传统机器学习		相同	相同
迁移学习	归纳式迁移	相同	相关
	无监督迁移	相关	相关
	直推式迁移	相关	相同

由于本文训练使用的数据库较为复杂，为了提高和优化模型的性能，实验利用迁移学习的方法来训练模型。首先在上百万的非人脸表情样本(来自 ImageNet 数据集图片分类数据)预训练本文的深度残差网络，利用人脸表情数据（FER-2013）对整个网络参数微调，得到表情识别网络，最后利用表情数据（CK+数据库）对网络进行训练，需要指出的是由于 CK+数据库数据较少，不能满足训练参数调节的需要，所以在最后的训练过程中保持特征提取层结构不变，仅对原表情识别网络中的全连接层与分类层的参数进行训练，使修改后的网络适用于人脸表情识别任务。在表情识别和图片分类等相关任务中，模型提取的特征具有良好的通用性，在相关任务中都能够取得良好的效果。因此，本文采用归纳式迁移学习方法。

3　实验结果与分析

为了验证算法较其他算法的特点，实验对比了本文算法与 Inception_V4、ResNet+softmax、VGG 四种网络的识别效果。识别效果如图 5 所示。

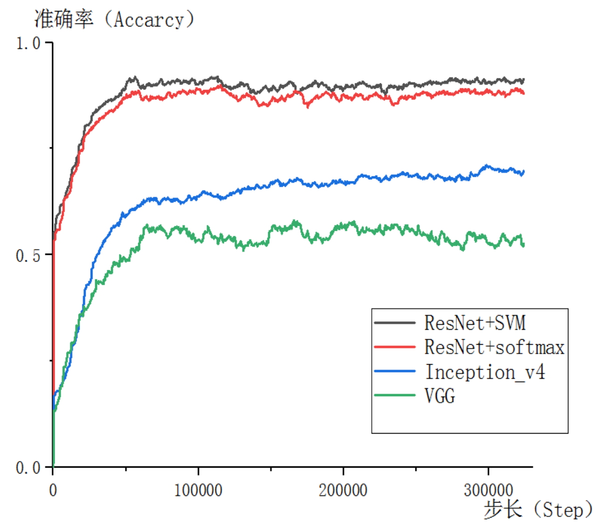


图 5　不同网络的识别效果

Fig. 5　Recognition accuracy of different networks

从图中可以看出 ResNet 网络要明显优于其他网络，使用 SVM 分类效果要优于 Softmax 的分类效果，提高了约 1%。

为了进一步评估实验结果，与其他方法进行了对比。表 2 显示了不同算法在 CK+和 GENKI-4K 的对比结果。

为了评估模型的泛化能力，用训练好的模型对 CGENKI-4K 笑脸表情数据库进行测试并与其他算法进行对比，如表 3 所示。

表 2　不同算法在 CK+数据库的对比结果

Table 2　The accuracies of different algorithms in CK+ database

算法	识别率（%）
CNN+AD[18]	84.55
CSPL+SVM[19]	89.89
LBP+CNN[20]	84.4
GB+DBNs+SAE[21]	92.46
LBP/VAR+DBN[22]	91.40
本文算法	91.33

表 3　不同算法在 GENKI-4K 上的对比结果

Table 3　Comparison of different algorithms on GENKI-4K

算法	识别率（%）
HOG+ELM[23]	88.50
LBP/Gray/Gabor+CRF[24]	91.14
Gabor+DAE[25]	90.75
本文算法	95.78

实验结果表明,本文提出的基于 ResNet+SVM 算法的识别率要优于传统的表情识别算法。较结合手工特征的深度学习算法，本文算法避免了复杂的显式特征提取，且识别率要优于部分深度网络。证明了该算法的可行性和有效性，在表情识别方面具有很好的泛化能力。

4　结束语

本文提出了 ResNet 结合 SVM 的算法解决了卷积神经网络的识别精度会随着网络深度下降这个问题，大大地增加了网络的深度，同时减少了网络的参数，这样不仅提高了网络的识别能力并且提高了网络的速度。SVM 的引入有效的提高了识别效率。不同于在表情识别中其他深度学习算法依赖人脸表情特征，本算法通过利用深度网络直接提取高阶特征。但本文算法亦有不足之处，本实验的重点工作集中在 RetNet 与 SVM 的结合上，并未对不同 SVM 的实验效果进行验证。未来的工作将集中在不同的 SVM 分类器对于实验的影响。

参考文献：

[1] LéCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Proc of International Conference on Neural Information Processing Systems. New York: Curran Associates Inc. 2012: 1097-1105.

[3] Taigman Y, Yang Ming, Ranzato M A, *et al.* DeepFace: closing the gap to human-level performance in face verification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2014: 1701-1708.

[4] Sun Yi, Wang Xiaogang, Tang Xiaoou. Deeply learned face representations are sparse, selective, and robust [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 2892-2900.

[5] He Kaming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 770-778.

[6] Notley S, Magdonismail M. Examining the Use of neural networks for feature extraction: a comparative analysis using deep learning, support vector machines, and K-nearest neighbor classifiers [EB/OL].

(2018-06-12) [2018-09-10]. https://arxiv.org/pdf/1805.02294.pdf.

[7] Agarap A F, Pepito F J H. Towards building an intelligent anti-malware system: a deep learning approach using support vector machine (SVM) for malware classification[EB/OL].. (2017-12-31)[2018-09-10]. https://arxiv.org/pdf/1801.00318.pdf.

[8] Kim S, Yu Zhibin, Kil R M, *et al.* Deep learning of support vector machines with class probability output networks [J]. Neural Networks, 2015, 64(4): 19-28.

[9] 时永刚，程坤，刘志文. 结合深度学习和支持向量机的海马子区图像分割 [J]. 中国图象图形学报, 2018, 23(4): 542-551. (Shi Yonggang, Cheng Kun, Liu Zhiwen. Segmentation of hippocampal subfields by using deep learning and support vector machine [J]. Journal of Image & Graphics, 2018, 23(4): 542-551.)

[10] Zhang Kaipeng, Zhang Zhanpeng, Li Zhifeng, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014, 52(3): 1-14.

[12] Chang C C, Lin C J. LIBSVM: a library for support vector machines [J]. ACM Trans on Intelligent Systems and Technology, 2011, 2(3): articleNo. 27.

[13] Russakovsky O, Jia Deng, Hao Su, *et al.* ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[14] Goodfellow I J, Erhan D, Luc C P, *et al.* Challenges in representation learning: a report on three machine learning contests. [J]. Neural Networks the Official Journal of the International Neural Network Society, 2013, 64: 59-63.

[15] Lucey P, Cohn J F, Kanade T, *et al.* The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2010: 94-101.

[16] Whitehill J, Littlewort G, Fasel I, *et al.* Toward practical smile detection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31(11): 2106-2111.

[17] Pan Jialin S, Yang Qiang. A survey on transfer learning [J]. IEEE Trans on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[18] 卢官明，何嘉利，闫静杰，等. 一种用于人脸表情识别的卷积神经网络 [J]. 南京邮电大学学报:自然科学版 ，2016, 36(1):16-22. (Lu Guanming, He Jiali, Yan Jingjie, *et al.* Convolutional neural network for facial expression recognition [J]. Journal of Nanjing University of Posts & Telecommunications, 2016, 36(1): 16-22.)

[19] Zhong Lin, Liu Qingshan, Yang Peng, *et al.* Learning Multiscale Active Facial Patches for Expression Analysis [J] IEEE Trans on Cybernetics, 2015, 45 (8): 1499-1510.

[20] 杨晓龙.基于局部特征提取和深度学习的人脸表情识别研究[D]. 重庆:重庆理工大学,2018.(Yang Xiaolong, Facial expression recognition based on local feature extraction and deep learning [D]. Chongqing: Chongqing University of Technology, 2018.

[21] 黄寿喜，邱卫根. 基于改进的深度信念网络的人脸表情识别 [J]. 计算机工程与设计, 2017, 38(6): 1580-1584. (Huang Shouxi, Qiu Weigen. Facial expression recognition via improved deep belief networks [J]. Computer Engineering & Design, 2017, 38(6): 1580-1584.)

[22] 何俊，蔡建峰，房灵芝，等. 基于 LBP/VAR 与 DBN 模型的人脸表情识别 [J]. 计算机应用研究，2016, 33(8): 2509-2513. (He Jun, Cai Jianfeng, Fang Lingzhi, *et al.* Facial expression recognition based on LBP/VAR and DBN model [J]. Application Research of Computers, 2016, 33 (8): 2509-2513.)

[23] An Le, Yang Songfan, Bhanu B. Efficient smile detection by Extreme Learning Machine [J]. Neurocomputing. 2015, 149 (PA): 354-363.

[24] 罗珍珍，陈靓影，刘乐元，等. 基于条件随机森林的非约束环境自然笑脸检测 [J]. 自动化学报，2018, 44(4): 696-706. (Luo Zhenzhen, Chen Jingying, Liu Leyuan, *et al.* Conditional Random Forests for Spontaneous Smile Detection in Unconstrained Environment [J]. Acta Automatica Sinica, 2016, 44(4): 696-706.)

[25] Liang Shufen, Liang Xiangqun, Guo Min. Smile recognition based on deep Auto-Encoders [C]//Proc of the 11th International Conference on Natural Computation. Piscataway, NJ: IEEE Press, 2016: 176-181.